

COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Dataset version: 2021-07-19

Total records in dataset: 27,185,885

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	27,185,885	27,185,885
total_cells	326,230,620	81,557,655
missing_fields	65,823,895	870,253
missing_pct	20%	1%
complete_fields	260,406,725	80,687,402
complete_pct	80%	99%
unknown_fields	22,643,984	9,566,067
unknown_pct	7%	12%
suppressed_fields	110	110
suppressed_pct	0%	0%
available_fields	237,762,631	71,121,225
available_pct	73%	87%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	15	0.0%	629,053	2.3%
sex	23	0.0%	52,227	0.2%
age_group	72	0.0%	188,973	0.7%
records_with_any_field	72	0.0%	811,779	3.0%